ERIC Document Resume	See Instruction Sheet	DATE OF RESUME	DATE RECEIVED

A. PUBLICATION DATA

1. ERIC NUMBER	2. ORIGINATOR CONTROL NO(S).	3. IDENTIFICATION NO.
ED002435	N.A.	2224

DO NOT WRITE IN SPACE BELOW

4. AUTHOR(S): Hirsch, Richard S.

5. TITLE OF DOCUMENT ABSTRACTED: The Effects of Knowledge of Test Results on Learning of Meaningful Material

6. PUBLICATION TITLE: N.A.

7. EDITOR(S): N.A.

8. PLACE OF ORIGIN ORG. The Instructional Film Research Program, The Pennsylvania ST. State College, Philadelphia, CITY Pennsylvania STATE

9. PUBLISHER OR SOURCE The Instructional Film Research Program, The Pennsylvania State College, Philadelphia, Pa.

10. VOL.	11. PAGES	12. MO./YR.	13. TOTAL PP.
N.A.	N.A.	9/52	29

TELE. EXT. INST/LOC CODE

MICROFILM COPY AVAILABLE YES ☒ NO ☐

COPYRIGHT GRANTED YES ☐ NO ☒

COPYRIGHT RELEASE FILED YES ☐ NO ☐

TYPE OF REPORT
PROPOSAL ☐
IN-PROGRESS ☐
FINAL ☒
FOLLOW-UP ☐

OTHER ERIC REFERENCES:

ERIC SATELLITE CODE

B. INDEXING DATA

14. SUBJECT AREA

15. TYPE OF ENDEAVOR Report

16. INDEXING TERMS (6-12)

knowledge of results	teaching methods	evaluation	
instructional film	test construction	achievement	
audiovisual aids	test results	retention	
	measurement		

17. ABSTRACT (MAX. 250 WORDS) The effect of knowledge of results on highly motivated subjects in a realistic training situation with highly meaningful and relevant learning material was investigated. The purpose of the study was to discover how knowledge of results could be applied to improve learning from films.

Subjects were midshipmen of the NROTC unit at the Pennsylvania State College. They were divided into matched groups in terms of semester standing in NROTC. Knowledge of results was defined as the presentation, in a number of ways, of information to a learner that his choice of answer to a test question was either correct or incorrect. Used in the study were two control and four experimental methods and six filmed presentations.

Knowledge of results facilitated learning. In most cases the more meaning that was explicitly given as the correct answer, the more it aided learning. It was better to provide the student with a statement of the question along with the answer than to give only the number or letter of the answer.

Military instructors improved their teaching by using tests to teach as well as to evaluate trainees. After giving the answer in the context of the question, the instructors showed an instructional film a second time. The showing aided retention of the material and reassured the trainee.

ERIC No.

18. SOURCE OF ABSTRACT

C. FUNDING DATA

19. SUPPORTING ORGANIZATION	20. DIVISION OR DEPT.	21. CONTRACT NO. / PROJECT NO.
	22. TOTAL COST OF PROJECT	
	23. DURATION OF PROJECT	U.S. Department of Health, Education, and Welfare Office of Education

ERIC Form 1

THE EFFECTS OF KNOWLEDGE OF TEST RESULTS ON LEARNING OF MEANINGFUL MATERIAL

Human Engineering Report SDC 269-7-30

Under Contract N6onr-269, T.O. VII Research Conducted by:
With The Pennsylvania State College Richard S. Hirsch
Instructional Film Research Program
State College, Pennsylvania
September 1952

- -

FOR THE PENNSYLVANIA STATE COLLEGE

Dean M. R. Trabue
Responsible Administrator

C. R. Carpenter
Program Director

- -

Approved:

For the Special Devices Center

T. B. Haley, CAPT., USN
Commanding Officer and Director

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education

This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

ED002435

SPECIAL DEVICES CENTER
OFFICE OF NAVAL RESEARCH
HUMAN ENGINEERING DIVISION

Problem:

It is a known fact that a person can learn from a testing experience when he is given the correct answers to the questions. In this way the test situation provides information to the student as well as the instructor. This research was done to discover how knowledge of results could be applied to improve learning from films.

Results:

This research showed that knowledge of results facilitates learning. In general, the more meaning that was explicitly given as the correct answer, the more it aided learning. When giving answers to questions it is best to provide the student with a statement of the question along with the answer rather than use a common procedure: "The answer to question 33 is D".

Recommendations:

The military instructor can therefore improve his instruction by using tests to teach as well as to evaluate trainees. The correct answer should be presented in statement form after each item whenever possible. Do not give just the number of the correct answer to the question but state the fact and place it in context. Give a second showing of an instructional film after the test; the showing will aid retention of the material and will reassure the trainee.

Loran C. Tunford Jr. Ph.D.
Project Engineer

J. Lanfield Daniels Ph.D.
Head, Applications Research Section

Clifford P. Seitz Ph.D.
Head, Program Branch

THE EFFECTS OF KNOWLEDGE OF TEST RESULTS ON LEARNING OF MEANINGFUL MATERIAL[1]

by

Richard S. Hirsch

The Instructional Film Research Program
The Pennsylvania State College

INTRODUCTION

Considerable research has been reported on the subject of knowledge of test results, but much of it has been concerned with human beings learning nonsense-material. The present study was designed to investigate the effect of knowledge of results on highly motivated subjects, in a realistic training situation, and with highly meaningful and relevant learning material.

Knowledge of results in this experiment was operationally defined to mean the presentation, in a number of ways, of information to a learner that his choice (from among five choices) of answer to a question in a test was either correct or incorrect. A number of methods of presenting the information was made possible by use of the Classroom Communicator. This study employed two control and four experimental methods and utilized the Classroom Communicator to provide the following:

1. A means for each subject to signal his answer to a question from among five choices.

2. A means to record the responses of the individuals so that the data could be collected and analyzed.

3. Finally, a means to inform each individual (a) whether or not his choice was the correct response, and (b) what the correct response should have been.

For a complete description of the Classroom Communicator and related equipment, the reader is referred to Technical Reports SDC 269-7-14 and SDC 269-7-15.

---

[1] This report is based on a dissertation by R. Hirsch submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, Stanford University, 1951.

Learning in this study was measured as retention. That is, it was measured as a difference obtained between two tests, the first test accompanied by knowledge of results and the second three weeks later without specific awareness of results. It is recognized that there are other definitions of learning; however, in this study learning was considered operationally, namely, as the retention of specific material.

EXPERIMENTAL DESIGN AND PROCEDURES

The Experimental Design

In order to measure the effect of each of the six different methods of presenting knowledge of results and to make comparisons among them, a Latin-square experimental design was employed. Basically, a Latin-square is an arrangement of Latin letters in rows and columns such that each letter occurs once, and only once, in each column and row.

Table 1 shows the Latin-square used in this experiment.

TABLE 1

LATIN-SQUARE DESIGN USED IN KNOWLEDGE OF RESULTS STUDY

(Capital letters represent methods)

	GROUPS					
FILMS	I	II	III	IV	V	VI
Aerodynamics	A	B	C	D	E	F
Carbine .30 Cal.	B	C	D	E	F	A
Guided Missiles	C	D	E	F	A	B
Hydrostatic Fuses	D	E	F	A	B	C
Plotting Board	E	F	A	B	C	D
Star Identification	F	A	B	C	D	E

The letters A, B, C, D, E, and F represent the independent variable; this variable was the six different methods involving knowledge of results. The rows and columns may also be considered as independent variables in a Latin-square. Thus, the design contains three variables which may be manipulated, the three variables being statistically independent of each other. In Table 1 the rows were assigned to the six films and the columns to the six groups of subjects.

In this arrangements, differences between the groups produce intercolumn variation; film differences produce interrow variations; and the methods of presenting knowledge of results (the experimental variables, designated by Latin letters) are orthogonal to each of the other two variables. Thus, the Latin-square provided a scheme of rotation such that each group of subjects was tested under each of the experimental conditions and was shown all the films. On the other hand, no subject was tested by any one method more than once nor was he shown any one film more than once.

The Experimental Variable

In this experiment the experimental variable is the method of presenting knowledge of results. The methods are as follows:

Method A: This was a control method; no knowledge of results was given in any manner.

Method B: Knowledge of results was presented by a neon lamp on each response station which lighted up for those subjects who had chosen the correct answer. It is to be noted that while the subject who gave an incorrect choice knew it because his response station light remained dark, he did not know which of the other four choices was the correct one.

Method C: This was the same as Method B, but with the addition that the Classroom Correct Answer Light now went on, informing all subjects, regardless of the choices made, what the number of the correct choice was. When the Classroom Correct Answer Light went on, the question was removed from the screen.

Method D: This method was also like Method B, but with the addition that the question was repeated on the screen with all alternatives removed except the correct one. This was done to determine the difference between (1) associating with the question a number of a correct answer (Method C) and (2) the correct answer itself.

Method E: This was the same as Method D, but with the addition that the end of the test was followed by a second showing of the film. The reason for a second film showing was that in this way the students would be able to check the test material and possibly tc see the rationale of each question.

Method F: This method was thought necessary in order to provide a control for Method E by showing the class the film, giving the test without knowledge of results, and then repeating the film a second time. The only difference between Methods E and F is that E provided knowledge of results.

Figure 1 illustrates the six methods employed in this experiment.

The problem for the experiment was to determine which of these six methods produced the greatest learning as measured by retention of specific material.

Experimental Procedure

Table 2 summarizes the entire experimental procedure for any one group.

TABLE 2

DETAILS OF METHODS USED FOR A SINGLE GROUP

PROCEDURES	METHODS					
	A	B	C	D	E	F
Pretest (30 items)	x	x	x	x	x	x
"Finger-exercise" on Classroom Communicator	x	x	x	x	x	x
Film screening	x	x	x	x	x	x
Immediate test on Classroom Communicator	x	x	x	x	x	x
Knowledge of results by:						
(1) Response station light		x	x	x	x	
(2) Classroom correct answer light			x*			
(3) Question and correct answer repeated				x	x	
Second film screening					x	x
Composite post test (3 weeks later)	x	x	x	x	x	x

* Note that this gave only the information about the number of the correct answer.

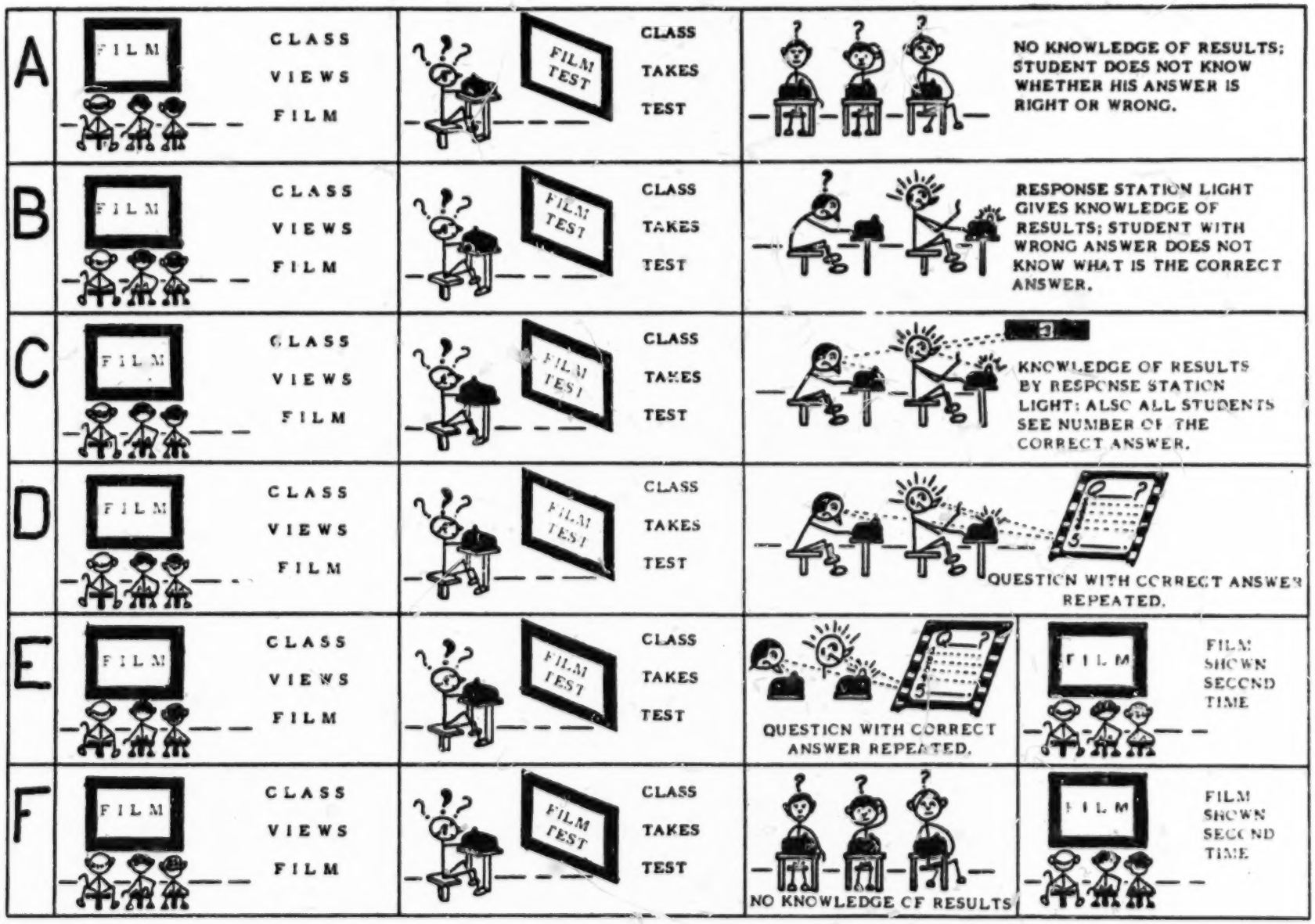


Fig. 1 - The six teaching methods.

The major steps in the experiment involved the following:

1. A pre-test was administered to obtain a measure of group homogeneity and of the pre-knowledge of the subject matter of the films. The pre-test consisted of thirty items, five from each of the six film-tests. It was administered to the NROTC midshipmen assembled as a single unit.

2. Each group met in a room where the Classroom Communicator was located. The classes met three times a week on alternate days for a period of two weeks making a total of six sessions. At each meeting, a group saw one of the six films, a different film and different method of presenting knowledge of results being used each time with a given group. At the beginning of each session, the subjects filled out attendance cards, from which seat numbers on the polygraph record were paired with the subjects' names.

(a) At the end of each film showing, a brief "finger exercise" on the Classroom Communicator was administered. This consisted of ten responses intended to refresh the subjects' memories about the finger-numbering. During this period the equipment was given an operational check.

(b) At the end of the "finger exercise," the controls on the console were set in accordance with requirements for the particular experimental method in effect, and a test based on the film was then projected on the screen, item by item. The subjects had a total of 35 seconds to read the question and answers, to choose one of the five alternatives, and to press the key corresponding to the choice made. The warning bell was set to ring at the end of 26 seconds, signalling that only 9 seconds remained of the 35 second response-period.

3. Depending on which method was in use during any particular session one or more of the following occurred on "lock stations".[2]

(a) If method A was the procedure, the item on the screen was advanced to the next question, and the response-period-switch was thrown to "response period" position.

(b) If method B was the procedure, the neon lamp would light on all response stations in which the correct response had been made. The question was removed from the screen. Thus, the subjects whose response station lamps remained dark were presumably uninformed concerning which of the alternatives was the correct one. After five seconds, a second question was projected on the screen and the procedure repeated.

[2] "Lock stations" is the operational condition of the equipment during which subjects cannot make a response or change a previous response.

(c) In method C, when the system went into "lock stations," the question was removed from the screen, the response station lamp came on (for stations with correct answers) and the number of the correct answer appeared in the Classroom Correct Answer Light Box. Thus, subjects whose responses were incorrect were informed of the number of the key they should have pressed. After five seconds the next question appeared on the screen and the process was repeated.

(d) When the system went into "lock stations" during method D, the question with its five alternatives was replaced on the screen by the question with only the correct answer listed. Thus, all subjects were able to check the accuracy of their responses, and also to see (in words) what the correct answer was to the question. After 10 seconds, the item on the screen was advanced, and the procedure was repeated.

(e) Method E was the same as method D, except that the last item on the test was followed immediately by a second showing of the film.

(f) Method F was almost the same as method A (both control groups), each item followed by the next without any form of knowledge of results. The only difference was that (as in method E) the last item was immediately followed by a second showing of the film.

4. The final step in the procedure was the administration of a 120-item composite post-test. This was accomplished three weeks after completion of the "immediate" tests with the entire NROTC unit assembled in one room.

The Films

Six United States Navy training films were used inthis experiment. The selection of films was such that the subject matter ranged from simple to difficult. This was done to determine the effect of knowledge of results on more than one type of film and on a range of difficulty. The films used were:

1. Aerodynamics - Theory of Flight
2. Carbine Cal. 30 M1 - Principles of Operation
3. Guided Missiles - Theory of Operation
4. Hydrostatic Fuses
5. Navigation - Plotting Board - The Mark 3A Plotting Board
6. Navigation - Star Identification

The Tests.

Table 3 summarizes the reliability of each of the tests based on the six films. It indicates that the tests were of moderate reliability. The number of items in each test was determined by the time available for testing in the longest of the six methods of presenting knowledge of results.

TABLE 3

THE RANGE OF RELIABILITY COEFFICIENTS FOR EACH FILM TEST

Film-test	Items	Alternate-form	K-R r_{tt}
Aerodynamics	21	.77 - .88	.58 - .78
Carbine	16	.68 - .88	.58 - .75
Guided Missiles	22	.81 - .88	.66 - .75
Hydrostatic Fuses	18	.71 - .78	.46 - .74
Plotting Board	21	.66 - .80	.47 - .80
Star Identification	22	.75 - .82	59 - .81

The six tests combined totalled 120 items. The post-test (delayed-recall test) consisted of the same 120 questions, the items of each test being randomly distributed throughout the post-test.

In addition to making up the post-test, five items from each of the six tests were randomly selected and administered to the midshipmen in advance of the experiment. The purpose of this pre-test was to establish a comparison between the groups in terms of previous knowledge of the material of the tests.

Each question (and its answers) was photographed on a 35mm. filmstrip for use during the experimental sessions (when knowledge of results was given). Figure 2 is a photograph showing sections from the pairs of one of the filmstrips. It is to be noted that one of a pair shows the same questions repeated with all but the correct answer deleted. This was used with methods D and E. The other filmstrip of the pair was used in testing all the other methods. As will be seen in the photograph, many of the test items were pictorial.

The Groups Studied

The experimental subjects were midshipmen of the NROTC unit at The Pennsylvania State College. These men were divided into groups on the basis of their time available for the experiment. Groups were matched, as nearly as possible, in terms of semester-standing in NROTC.

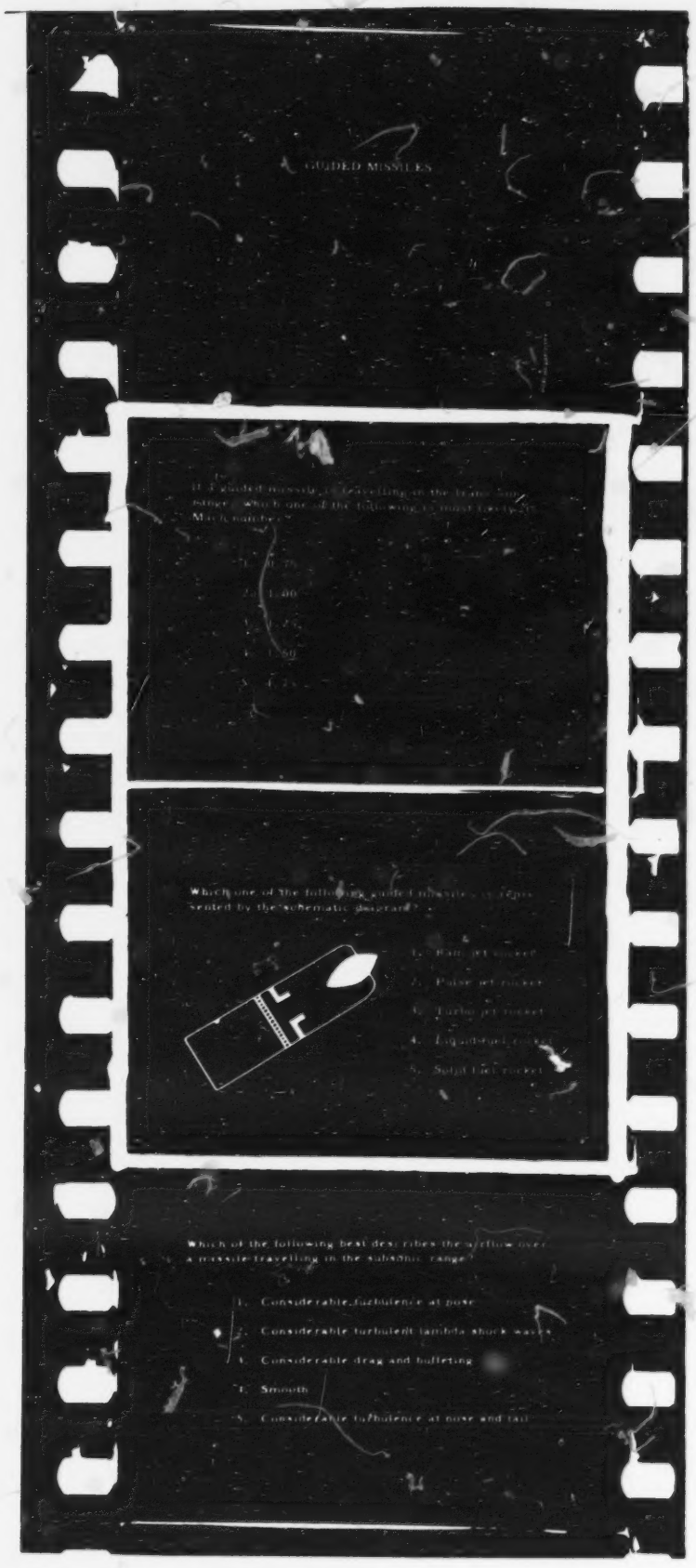
In the final analysis, after loss in population through absences, the smallest group had 23 men. To conform to Latin-square requirements, the other five groups were also reduced by random selection, to 23 men, a total of 138 men in the six groups.

RESULTS

Preliminary Test

Table 4 presents the analysis of variance of the scores on the pre-test.

For this test the ratio of (1) variance between groups to (2) variance within group is 1.30, and this value is well below the 5 per cent value of 2.29 (for 5 and 132 degrees of freedom) required for significance. Accordingly, within the limits of the analysis, the groups appear to have been homogeneous at the start of the experiment.


A. GUIDED MISSILES
Which one of the following guided missiles is repre-
sented by the schematic diagram?
1. Ram jet rocket
2. Pulse jet rocket
3. Turbo jet rocket
4. Liquid fuel rocket
5. Solid fuel rocket
Which of the following best describes the airflow over
a missile travelling in the subsonic range?
1. Considerable turbulence at nose
2. Considerable turbulent lambda shock waves
3. Considerable drag and buffeting
4. Smooth
5. Considerable turbulence at nose and tail

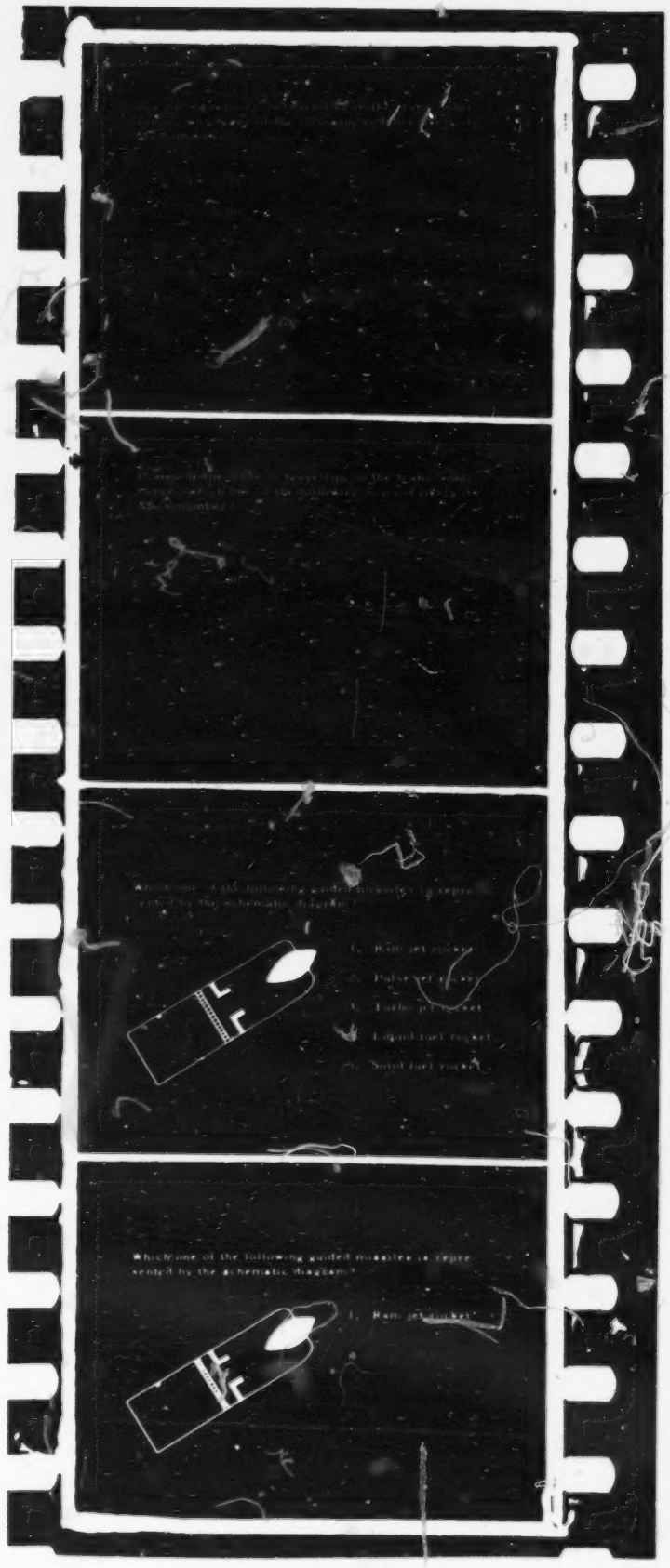


Fig. 2 - Photographs of section of filmstrip test on Guided Missiles. The right hand photograph illustrates one method of presenting knowledge of results.

TABLE 4

ANALYSIS OF VARIANCE OF SCORES ON PRELIMINARY TEST (30 Items)

Source of Variance	Degrees of Freedom	Sum of Squares	Mean Square	F
Between groups	5	54.32	10.86	1.30
Within groups	132	1100.26	8.34	....
Total	137	1154.58	.....	.....

Immediate Tests.

The homogeneity between groups found in the analysis of the preliminary test appears to break down in subsequent tests. However, the design of the experiment was such that every group underwent every method and saw every film; hence, it was judged that the design would tend to minimize, if not eliminate, the effects due to individual differences.

Table 5 summarizes the analysis of variance of the immediate tests. From this it may be seen that the variance due to differences between groups on the six film-tests is highly significant. Moreover, it appears

TABLE 5

ANALYSIS OF VARIANCE OF THE IMMEDIATE TEST

Source of Variance	D. F.	Sums of Squares	Mean Square	F
Independent Observations:				
Between groups	5	440.50	88.10	4.96
Residual between individuals (error)	132	2345.01	17.77	....
Total between individuals	137	2785.51	.....	....
Correlated Observations:				
Films	5	5642.53	1123.51	201.70
Treatments	5	73.27	14.65	2.62
Latin-squares	35	6284.55	.....	....
Residual Latin-squares (error)	20	128.24	6.41	1.15
*Residual within individuals (error)	660	3676.29	5.57	....
Combined error	680	3804.53	5.59	....
Total within individuals	690	9520.33	.....	....
Total for experiment	827	12305.85	.....	....

* Tested for homogeneity of variance by Bartlett's test; null hypothesis acceptable and data was pooled.

that the difference attributable to methods is also significant (just beyond the 5 per cent level of confidence). A closer analysis of the data (Table 8 in the Appendix) indicates that the obtained difference is due to but one test and group, and that, except for this one difference, none of the other methods shows any statistical significance on the immediate test. The lack of method-differences (except for the one case noted) can be accounted for by the fact that the learning material was comprised of relatively discrete items, and there was no opportunity for intra-serial effects to operate.

In view of the fact that the significance of the variance among methods is due almost entirely to the scores of one group on a particular film, it is considered reasonable to say that there is no significant difference between the methods on the immediate test.

The difference between the groups, however has been noted as highly significant. This may be due to the fact that some groups, having a higher percentage of upper-classmen than others, had a greater background of Naval course work to fall back on when they took the immediate tests. It is possible also that the difference in the hours during the day when the groups met may have had an effect on their performance. This was, of course, the one experimental factor which was not controlled, and its effect remains unknown.

The Delayed Post-Test.

The results of the delayed post-test indicate that both the film and group variances were considerable and were statistically significant. Moreover, the results show that the variance due to methods-differences is also significant. Figure 3 presents a graph of the average means for each method. Meaningful as this data is, the comparison between the immediate test and the delayed post-test provides more revealing information.

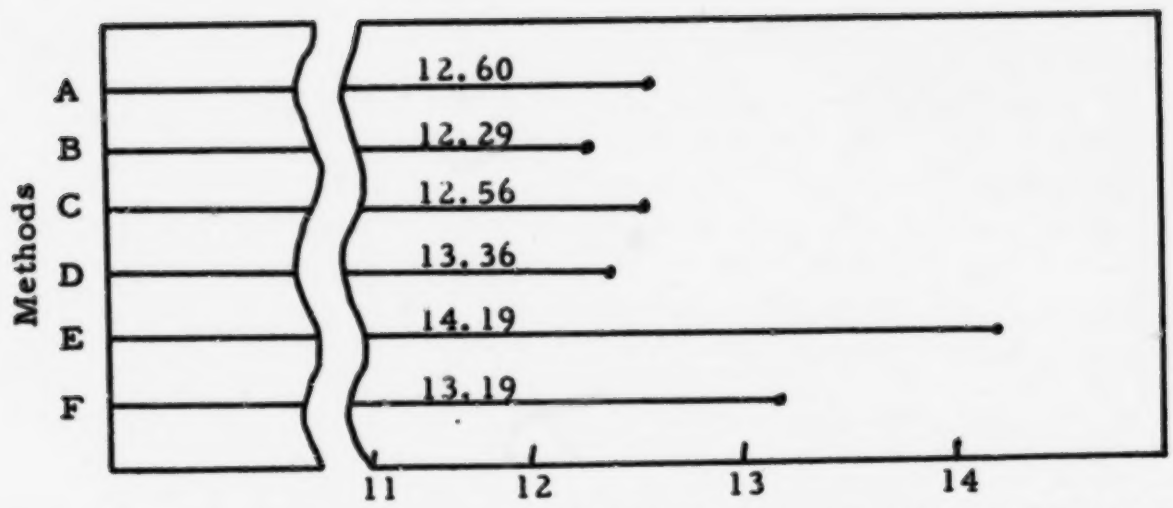


Fig. 3 - Average Mean scores of the Methods on the Delayed Post-Tests. The break emphasize the difference in Means

The Differences: Post-Test minus Immediate Test

Table 6 presents the analysis of variance of the difference-scores between the delayed post-test and immediate test.

The data indicates that the variance due to groups is significant at the 5.0 per cent but not at the 1.0 per cent level.

A test for the homogeneity of the within-individuals indicated that the six estimates could be pooled. However, the F-ratio for the comparison between the residual-within-subjects and residual-Latin-squares indicated that the two error variances were not homogeneous; hence, the F-ratios are given in terms of each error variance. That is the error terms in the analysis of variance (residual Latin-squares and residual within individuals) were significantly different from each other ($F = 2.21$) at the .01 level; hence, each of the error-mean squares was used to derive the F-ratios for the methods. Using the residual Latin-squares error, the F-ratio (6.47) for methods is well beyond the .01 level of significance; using the residual-within-individuals error value, the methods are significant beyond the .001 level ($F = 14.32$).

TABLE 6

ANALYSIS OF VARIANCE OF THE DIFFERENCES (DELAYED POST-TEST MINUS IMMEDIATE TEST)

Source of Variance	D.F.	Sum of Squares	Mean Square	F
Independent Observations:				
Between groups	5	102.89	20.58	2.61*
Residual between individuals (error)	132	1041.49	7.89	....
Total between individuals	137	1144.39	.....	....

				F	
				RWI	RLS
Correlated Observations:					
Films	5	230.76	46.15	8.82	3.99
Treatments	5	374.53	74.91	14.32***	6.47**
Latin-squares	35	939.50	.....	.....	....
Residual Latin-squares (error)	20	231.41	11.57	2.21**	....
Residual within individuals (error)	660	3453.46	5.23	.....	....
Total within individuals	690	4290.17	.....	.....	....
Total for experiment	827	5434.55	.....	.....	....

* .05 level significance
** .01 level significance
*** .001 level significance

Accordingly, the significance of the differences indicated between the methods on the delayed post-test becomes clearer when analyzed in terms of the difference scores.[3]

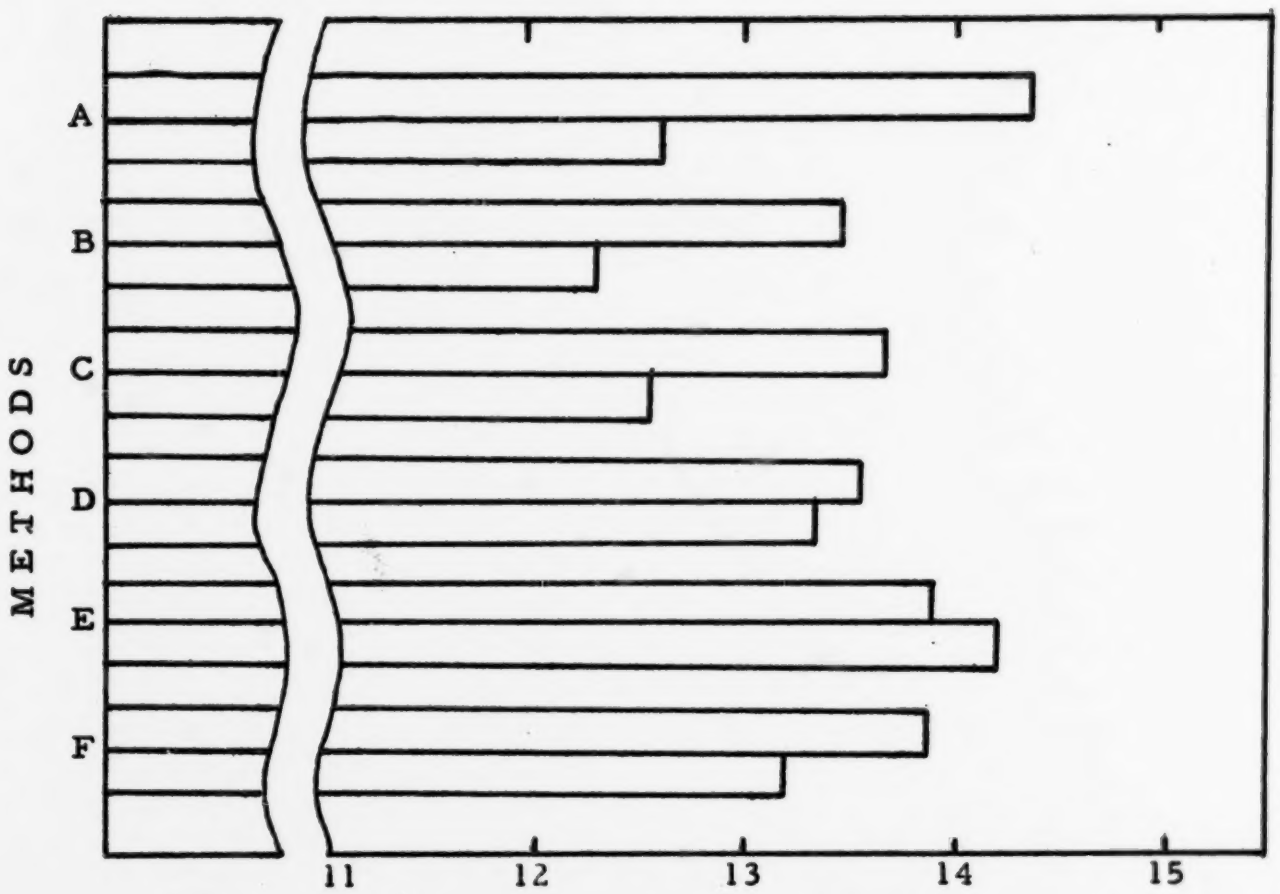


Fig. 4 - Comparison of the Delayed Post-Test and Immediate Test for the Six Methods. The break is intended to emphasize the differences. The upper bar represents the immediate test; the lower bar is the post-test.

Table 9 (appendix) lists the difference means, difference-standard deviations and t-ratios for each combination of film and method. This table indicates that the expected forgetting between the two tests did occur in most cases. In Method A all the groups showed a considerable loss in retention. The same is true for Method B. In Method C, however, Group V scored higher on the delayed post-test on the Plotting Board than they scored on the immediate test.

---

3 "Difference" scores are not ordinarily likely to be highly reliable. However, when it can be shown, as in the present case, that significant differences can be found, the problem of reliability is of secondary importance.

Group VI had the combination of Plotting Board and Method D, and no loss of retention occurred. Method E, it will be noted, resulted in an increase of learning. That is, in combination with four films (Aerodynamics, Carbine, Hydrostatic Fuses, and Star Identification) Method E produced scores on the delayed post-test which were higher than those earned on the immediate test. Figure 4 above summarizes the data.

Interpretations and Conclusions

It may be seen from the highly significant t-ratios (Table 9 in the Appendix) that Methods A, B, C, and F all have produced considerable and significant loss in retention between the immediate tests and the delayed post-test. That is, each loss is significantly greater than a zero-loss. Method D, on the other hand, also had some loss, but the loss was held to a minimum. The t-ratio for Method D indicates that the loss is not significantly greater than that which might be expected to occur by chance. Fig. 5 presents the extent and direction of the differences between these average means.

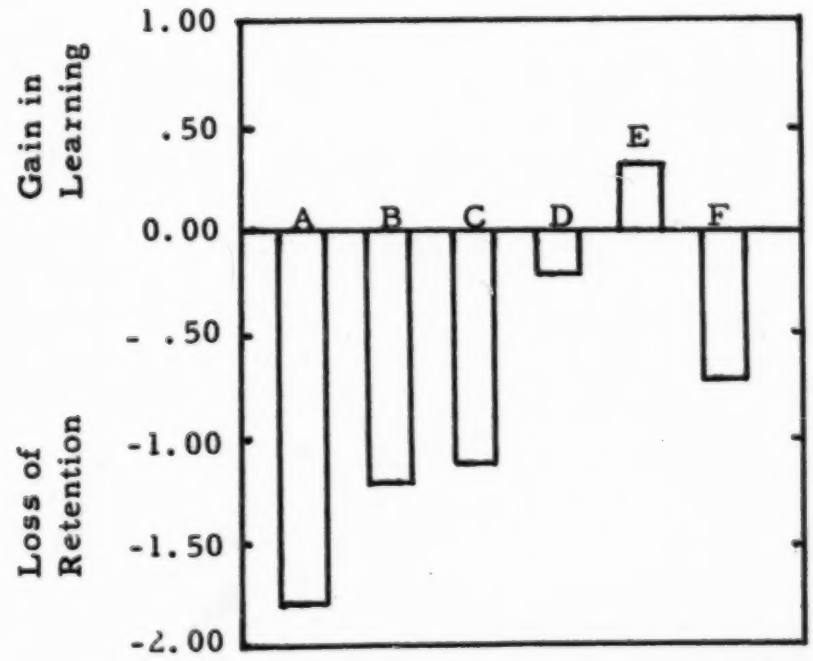


Fig. 5 - Showing the extent and direction of the differences between the average means (Delayed post-test minus the Immediate test). Note that while methods A, B, C, D and F have undergone a loss of learning, method E has gained.

When Method E is considered, however, it can be seen that, although the learning gain between the delayed post-test and the immediate test was not great (+0.29) and was not statistically significant, it is the only method which resulted in a learning gain (not a retention loss). Moreover, of the six films, four resulted in a gain under Method E.

Table 7 summarizes the differences among methods in terms of retention losses (delayed post-test minus immediate test) for the six methods. In this table, the difference-mean for each method has been subtracted from every other difference-mean, resulting in a series of difference-differences.

TABLE 7

DIFFERENCES AMONG METHODS RETENTION LOSSES

(DELAYED POST-TEST MINUS IMMEDIATE TEST) FOR THE SIX METHODS

METHODS		A	B	C	D	E	F
	MEANS	-1.77	-1.17	-1.12	-0.21	+0.29	-0.70
A	-1.77	X	-0.60*	-0.65*	-1.56**	2.06**	-1.07**
B	-1.17		X	-0.05	-0.96**	1.46**	0.47
C	-1.12			X	0.91**	1.41**	0.42
D	-0.21				X	0.50**	0.49
E	+0.29					X	0.99**
F	-0.70						X

* d at .05 = 0.54
** d at .01 = 0.72

NOTE: In every case, the Mean at the top of the column has been subtracted from the Mean at the side.

Calculation for a significant difference indicated that a difference of 0.54 was necessary to be significant at the .05 level; at the .01 level a difference of 0.72 was required.

Thus, Method A (Table 7) is significantly different from every other method. Methods B, C, and F are not significantly different from each other. Another cluster (i.e., non-significance between values) is formed by methods D and F. A third cluster is formed by Methods D and E.

The next logical step is the interpretation of the relative effectiveness of the methods.

Effectiveness of the Various Methods

The present experiment was undertaken to investigate the relative effectiveness of several methods of presenting knowledge of results on tests based on the content of sound motion pictures; hence, the most important findings are those which deal with the methods used. While an examination of the effect of the different film-tests and the different groups is possible and interesting, such a study would not be entirely relevant.

It is clear that the methods stand in a definite order of effectiveness, and it is proposed to discuss the significance of the methods in the order of effectiveness found in the results of the differences between the immediate test and the delayed post-test.

(1) The superiority of Method E.

In all methods except Method E it was found that there was considerable loss in retention of the material presented in the films as shown by the differences between the immediate and post-tests. Moreover, the retention-losses in most cases were very highly significant. Method E, on the other hand, not only was most effective in holding retention-losses to a minimum, but for some film tests it actually brought about a learning gain. For example, in combination with the film on Aerodynamics, the gain was 0.61. The next best method (D) also showed a gain on this film, but this was very small (0.04). Method E was the only one to show a gain (0.65) on the Carbine film-test. In combination with Guided Missiles, Method E was not as consistent as in its effect on other tests; the retention-loss was greater than that found under Methods D and F (in that order). With Hydrostatic Fuses, Method E was again the only one to result in a learning gain. The results on the Plotting Board, on the other hand, are not clear. Method E was not the best method. On the contrary, while Method E resulted in a loss of -0.83, Method D held the loss to zero, and Method C (with Group V) brought about a gain of +0.17. It will be recalled that it was Group V which had high scores under Method A (with Guided Missiles) on the immediate test (cf. Table 8 in the Appendix).

In combination with Star Identification, Method E resulted in a small gain (+0.04) while Method D (Group V again) showed a greater gain (+0.39).

Thus it may be seen that although Method E appears to be the most effective, it is somewhat inconsistent when considered film-by-film. This inconsistency, moreover, appears to be due in large part to the overall superiority of Group V. However, these findings do not necessarily invalidate the generalization that, overall, Method E is superior to the other methods. Group V appears only once under Method E (Aerodynamics), and it does contribute to the superiority of the method; nevertheless, the high-scoring Group V in other film-method combinations tends to detract from the superiority of Method E. Therefore, in spite of the performance of Group V, it is still possible to say of Method E that it is superior to other methods.

Accordingly, from these findings, it appears that giving the subjects knowledge of results by repeating the question along with its correct answer (in words) plus a second showing of the film (thereby permitting the subjects to check the rationale of the question and its answer) not only retards loss of retention, but also appears to consolidate the learning and to increase it somewhat.

(2) Method D.

Method D presented knowledge of results in exactly the same manner as Method E, namely, by repeating the question and its answer. The experimental difference between Methods E and D was that in the latter the film was not shown a second time.

Compared with Method E in effectiveness, the findings indicate that Method D is slightly, though not significantly, less effective in retarding forgetting than Method E. While Method E showed a slight gain in learning, Method D resulted in some slight loss of retention. However, the t-ratio, 1.03, indicates that Method D was able to hold the loss to a degree of statistical non-significance.

On two of the film-tests, Aerodynamics and Star Identification, Method D resulted in a positive gain in learning. On a third, Plotting Board, the loss from the immediate test to the post-test was held to zero.

(3) Method F

Method F stands third in the order of effectiveness. Although it is statistically significant in difference from Method E, it is not significantly different from Methods D, C, and B. This method consisted of a presentation of the film, administration of the test without knowledge of results (control), and a second showing of the film. In keeping with other studies on the subject, this method adds evidence to the finding that two showings of a film are superior to one showing, since Method F is significantly better than Method A (one showing of the film and test without knowledge of results).

Although the retention-loss was held to a point of non-significance by Method D and E (the latter showing a learning gain), Method F, on the other hand, was unable to keep the forgetting minimal ($t = 3.46$). At the other end of the scale, forgetting was greatest under Method A ($t = 9.89$). The t-ratios indicate that the loss in retention is significantly greater than zero, that is, zero loss of retention.

Perhaps the most important finding concerning Method F is that it is slightly better (though not significantly so) than C and B. Thus, within the limits of the experiment, it would seem that providing two showings of a film is at least as good a method for retarding forgetting as are those methods (B and C) which give knowledge of results by merely telling students who made a correct choice that they were right (by response station light) or by associating the number of the answer with a question.

(4) Method C

In Method C the Classroom Correct Answer Box lit up the number of the correct choice. If we continue to use a method's ability to hold forgetting to a minimum as a criterion of effectiveness, it may be seen that Method C is more effective than Method A (at the .05 level). Thus, it may be said that giving subjects the number of a correct answer immediately after they have made a choice is better than no knowledge of results at all.

On the other hand, it is interesting to note that there is practically no difference between the means of Methods C and B. This would seem to imply that merely telling those students who chose correctly that they have been right is as good (or is no less effective in minimizing forgetting) as that method which tells the students what the number of the correct answer was.

Comparing Method C with methods which are more effective, it can be seen that Method D (which associates the question with the answer in words) is significantly better (at the .001 level) than Method C (which associates the correct answer with a number). However, it should be noted that, when the number of the answer was given, the questions and answers were removed from the screen. This was an experimental technique which would not be used in actual practice with the Classroom Correct Answer Light. It should be noted that the equipment was designed to permit the use of the Classroom Correct Answer Light with the alternatives to a question remaining on the screen.

(5) Method B.

Method B was that method in which the response-station lamp was lighted for all subjects whose choice of answer was a correct one. Students whose answer was incorrect were not informed of what the correct answer should have been. Of all the methods in which knowledge of results is provided, Method B is least effective in minimizing the amount of forgetting. However, it is significantly better (between the .05 and .01 level) than no knowledge of results.

(6) Method A.

The inferiority of Method A relative to all the others has been adequately demonstrated in the previous discussions. With zero-forgetting as a criterion, the t-ratio of 9.89 indicates that, overall, the amount of forgetting was highly significant, and the inference is that giving a test without knowledge of results (although each subject was informed of his total scores at the end of the following meeting), will not retard forgetting very much, if at all. The extent to which this is different from the performance of a group which did not take the immediate test at all is unknown; in other words, how much learning occurred from the test administration alone is not known.

Method B, the next least effective method, with a t-ratio of 5.33 is significantly more effective, in spite of the fact that in Method B the only knowledge of results was informing a student that he had made a correct choice.

In summary, the following general statements may be made:

(1) Knowledge of test results, given item by item, aids retention of learned material.

(2) Knowledge of test results which takes the form of a repetition of each question with its correct answer, given immediately after a choice has been made, aids retention of the learned material almost to the point where little or no forgetting occurs (within, at least, the period of time covered by this experiment).

(3) When, in addition to the knowledge of results described in (2) above, a film is also repeated a second time, not only is retention aided, but also some learning gain occurs.

IMPLICATIONS

The results of this study suggest at least three generalizations of practical importance: (1) Tests can be used to teach as well as to evaluate trainees, and the instructor who uses tests only for evaluation deprives himself of an effective teaching aid; (2) Teaching methods which provide item-by-item knowledge of results still further facilitates learning and retention; (3) the effect of knowledge of results will be enhanced if the meaning of a choice is fully represented in words rather than by signs (in the form of numerals) corresponding to the choices in a multiple-choice examination.

These general statements have important implications in a mass-training program. The training of large numbers of individuals in a minimum time calls for the techniques which have been demonstrated to facilitate learning. The utilization of tests with provisions for knowledge of results is one such technique.

Still another implication emerges from this study; namely, the Classroom Communicator, used as a teaching aid, helps to minimize the unidirectionality (from lecturer to student) which usually prevails in most mass training situations. The Classroom Communicator provides two-way communication between the instructor and a number of students at the same time. In a sense, one instructor is able to hear the "recitation" of an entire class at once. In other words, the system provides for the active participation of large numbers of persons in the learning process.

Finally, this study adds still further support to the growing body of evidence that learning occurs from films. Moreover, the experimental method which utilized two showings of the films indicates, by the obtained results, that repeated screenings materially increase the learning.

SUMMARY

The purpose of this experiment was to determine whether knowledge of results, provided by means of the Classroom Communicator, would affect the learning of material from tests based on the contents of selected sound films.

Utilizing techniques available by use of the Classroom Communicator, four methods of presenting knowledge of results were employed in addition to two controls. The six methods were the following:

1. Small individual lamps were used to signal success to each subject if he chose the correct answer to a question.

2. The number of the correct choice was given to all subjects regardless of failure or success in choosing the answer.

3. The question was repeated with all but the correct answer deleted. Thus, this method is a more meaningful presentation of information than (2).

4. The film was presented a second time (after the test) in addition to knowledge of results by method (3) above.

5. No knowledge of results was presented in any manner. This was the first control method.

6. No knowledge of results was presented, but the film was presented a second time after the test. This was the second control method.

The subjects were 138 NROTC midshipmen divided into six groups of 23 men each. The tests were on material contained in six Navy training films. The experimental procedure utilized a Latin-square design, rotating the six groups, the six methods, and the six tests (films). An analysis of variance was applied to the obtained data.

The results led to the following conclusions:

1. Knowledge of results facilitates learning, since even information by means of signs (lights and numbers) is significantly better than no information at all.

2. The more meaningful the information about results on test questions (repetition of the question and correct answer with a second film showing), the more effectively it aids learning.

APPENDIX

TABLE 8

MEANS AND STANDARD DEVIATIONS ON IMMEDIATE TEST CLASSIFIED BY FILMS AND TREATMENTS

FILMS	METHODS					
	A	B	C	D	E	F
Aerodynamics	I	II	III	IV	V	VI
Means	16.91	16.35	15.13	15.09	16.74	17.78
S. D.	2.67	2.99	2.64	3.02	2.42	1.98
n=21*						
Carbine .30 Cal.	VI	I	II	III	IV	V
Means	12.39	11.26	10.78	9.30	11.39	11.65
S. D.	1.93	1.51	2.19	2.35	2.06	1.60
n=16*						
Guided Missiles	V	VI	I	II	III	IV
Means	18.97	17.17	17.43	17.65	16.26	17.04
S. D.	3.34	2.07	3.42	2.81	3.29	2.39
n=22*						
Hydrostatic Fuses	IV	V	VI	I	II	III
Means	11.52	11.35	12.17	12.22	12.00	10.13
S. D.	3.02	2.22	2.39	2.54	2.47	2.52
n=18*						
Plotting Board	III	IV	V	VI	I	II
Mean	13.74	14.83	16.26	16.13	15.61	15.04
S. D.	3.10	2.68	2.77	2.44	3.03	2.93
n=21*						
Star Identification	II	III	IV	V	VI	I
Mean	12.70	9.78	10.26	11.00	11.39	11.70
S. D.	3.03	3.37	2.54	3.83	2.78	3.00
n=22*						
Mean for method	14.37	13.46	13.67	13.57	13.90	13.89
S. D. for method	3.78	3.93	3.84	4.12	3.57	3.80
S. E. of method mean	.32	.34	.33	.35	.30	.32

* "n" refers to number of items in each test.

TABLE 9

MEANS, STANDARD DEVIATIONS, AND t-RATIOS
OF DIFFERENCES (POST-TEST MINUS IMMEDIATE TEST)
CLASSIFIED BY FILMS AND TREATMENTS

FILMS	METHODS					
	A	B	C	D	E	F
Aerodynamics	I*	II	III	IV	V	VI
Mean	-1.57	-1.04	- .91	+ .04	+ .61	- .43
S. D.	1.50	2.68	2.32	2.54	1.76	2.02
t**	4.90	1.83	1.85	0.08	1.62	1.01
Carbine .30 Cal.	VI	I	II	III	IV	V
Mean	-1.22	- .70	- .30	- .13	+ .65	- .13
S. D.	1.93	1.94	1.83	2.01	1.71	2.25
t	2.95	1.68	.78	.30	1.79	.27
Guided Missiles	V	VI	I	II	III	IV
Mean	-2.00	-1.87	-2.74	-1.04	-1.65	-1.30
S. D.	1.69	2.42	2.91	1.49	1.83	2.11
t	5.54	3.62	4.42	3.29	4.23	2.89
Hydrostatic Fuses	IV	V	VI	I	II	III
Mean	-1.83	- .70	-1.43	- .52	+2.91	- .48
S. D.	2.20	2.56	2.18	2.55	2.12	2.10
t	3.89	1.27	3.08	.95	6.43	1.07
Plotting Board	III	IV	V	VI	I	II
Mean	-1.52	- .13	+ .17	.00	- .83	- .87
S. D.	2.24	2.49	1.79	1.84	2.79	2.86
t	3.18	.25	.46	0.00	1.39	1.42
Star Identification	II	III	IV	V	VI	I
Mean	-2.48	-2.57	-1.48	+ .39	+ .04	-1.00
S. D.	2.57	2.41	3.03	3.19	3.59	2.60
t	4.53	4.99	2.29	.58	.06	1.80
Method Mean (difference)	-1.77	-1.17	-1.12	- .21	+ .29	- .70
Method S. D.	2.09	2.56	2.57	2.38	2.79	2.38
t	9.89	5.33	5.09	1.03	1.21	3.46

* Roman numerals refer to groups of men.

** "t" significance levels are as follows:

t	Significance level
1.72	.10
2.07	.05
2.51	.02
2.82	.01

TABLE 10

SUMMARY OF FORMULAS USED IN THE ANALYSIS OF VARIANCE

Source of Variance	Formula	D.F.
Independent Observations:		
Between Groups:	$\frac{\Sigma(\Sigma X_g)^2}{138} - C^*$	5
Residual between individuals (error)	I - G	132
Total Between Individuals	$\frac{\Sigma(\Sigma X_i)^2}{6} - C$	137
Correlated Observations:		
Films	$\frac{\Sigma(\Sigma X_f)^2}{138} - C$	5
Treatments:	$\frac{\Sigma(\Sigma X_t)^2}{138} - C$	5
Latin-squares:	$\frac{\Sigma(\Sigma X_{fg})^2}{23} - C$	(35)
Residual Latin-squares (error):	LS - F - T - G	20
Residual within individuals (error):	WI - F - T - RLS	660
Total for Experiment:	$\Sigma X^2 - C$	827

* $C = (\Sigma X)^2 / 828$